THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Binaural Rendering of Spherical Microphone Array Signals

A Real-Time Implementation to Evaluate the Effects of Additive Noise

Hannes Helmholz

Department of Architecture and Civil Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2021

Binaural Rendering of Spherical Microphone Array Signals
A Real-Time Implementation to Evaluate the Effects of Additive Noise

HANNES HELMHOLZ

Technical report: 2021:9
Lic / Architecture and Civil Engineering / Chalmers University of Technology

Department of Architecture and Civil Engineering
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000

Cover:
Visual representation of the first real spherical harmonics up to 3rd order. The individual modes are scaled independently for better illustration.

Printed by Chalmers Reproservice
Gothenburg, Sweden, November 2021

Binaural Rendering of Spherical Microphone Array Signals
A Real-Time Implementation to Evaluate the Effects of Additive Noise
HANNES HELMHOLZ
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

The presentation of extended reality for consumer and professional applications requires major advancements in the capture and reproduction of its auditory component to provide a plausible listening experience. A spatial representation of the acoustic environment needs to be considered to allow for movement within or an interaction with the augmented or virtual reality. This thesis focuses on the application of capturing a real-world acoustic environment by means of a spherical microphone array with the subsequent head-tracked binaural reproduction to a single listener via headphones. The introduction establishes the fundamental concepts and relevant terminology for non-experts of the field. Furthermore, the specific challenges of the method due to spatial undersampling of the sound field as well as physical limitations and imperfections of the microphone array are presented to the reader. The first objective of this thesis was to develop a software in the Python programming language, which is capable of performing all required computations for the acoustic rendering of the captured signals in real-time. The implemented processing pipeline was made publicly available under an open-source license. Secondly, specific parameters of the microphone array hardware as well as the rendering software that are required for a perceptually high reproduction quality have been identified and investigated by means of multiple user studies. Lastly, the results provide insights into how unwanted additive noise components in the captured microphone signals from different spherical array configurations contribute to the reproduced ear signals.

Keywords: Spherical microphone arrays, binaural rendering, real-time signal processing, additive noise.

List of Publications

This thesis is based on the following publications:

[A] **Hannes Helmholz**, Carl Andersson, Jens Ahrens, “Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals”. in Fortschritte der Akustik – DAGA 2019, Rostock, Germany, Mar. 2019, pp. 1462-1465.

[B] **Hannes Helmholz**, Tim Lübeck, Jens Ahrens, Sebastià V. Amengual Garí, David Lou Alon, Ravish Mehra, “Updates on the Real-Time Spherical Array Renderer (ReTiSAR)”. in Fortschritte der Akustik – DAGA 2020, Hannover, Germany, Mar. 2020, pp. 1169–1172.

[C] **Hannes Helmholz**, Jens Ahrens, David Lou Alon, Sebastià V. Amengual Garí, Ravish Mehra, “Evaluation of Sensor Self-Noise In Binaural Rendering of Spherical Microphone Array Signals”. in Proceedings of International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, May 2020, pp. 161-165, DOI: 10.1109/ICASSP40776.2020.9054434.

[D] **Hannes Helmholz**, David Lou Alon, Sebastià V. Amengual Garí, Jens Ahrens, “Instrumental Evaluation of Sensor Self-Noise in Binaural Rendering of Spherical Microphone Array Signals”. in Proceedings of Forum Acusticum, Lyon, France, Dec. 2020, pp. 1349–1356, DOI: 10.48465/fa.2020.0074.

[E] **Hannes Helmholz**, David Lou Alon, Sebastià V. Amengual Garí, Jens Ahrens, “Effects of Additive Noise in Binaural Rendering of Spherical Microphone Array Signals”. in IEEE/ACM Transactions on Audio, Speech, and Language Processing *(accepted for publication)*.

Contributions: The included papers were prepared in collaboration with the co-authors. The author of this thesis was responsible for the major progress of the work including taking part in the planning of the papers, designing and developing the software implementations, planning and conducting the perceptual user studies and writing the papers.

Other publications by the author, not included in this thesis, are:

[F] J. Ahrens, **H. Helmholz**, C. Andersson, “Authentic Auralization of Acoustic Spaces based on Spherical Microphone Array Recordings”. *Auditorium Acoustics*, Hamburg, Germany, Oct. 2018, pp. 1–8.

[G] T. Lübeck, J. M. Arend, **H. Helmholz**, J. Ahrens, C. Pörschmann, “Comparison of Mitigation Approaches of Spatial Undersampling Artifacts in Spherical Microphone Array Data Auralizations”. *Fortschritte der Akustik – DAGA 2020*, Hannover, Germany, Mar. 2020, pp. 623–627.

[H] T. Lübeck, **H. Helmholz**, J. M. Arend, C. Pörschmann, J. Ahrens, “Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data”. *Journal of the Audio Engineering Society, vol. 68, no. 6*, June 2020, pp. 428–440, DOI: 10.17743/jaes.2020.0038.

[I] T. Lübeck, J. M. Arend, C. Pörschmann, **H. Helmholz**, J. Ahrens, “Perceptual Evaluation of Mitigation Approaches of Impairments Due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data: Dry Acoustic Environments”. *23rd International Conference on Digital Audio Effects*, Vienna, Austria, Sep. 2020, pp. 250–257.

[J] J. Ahrens, **H. Helmholz**, D. L. Alon, and S. V. Amengual Garí, “The Far-Field Equatorial Array for Binaural Rendering”. *International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021, pp. 421–425, DOI: 10.1109/ICASSP39728.2021.9414368.

[K] J. Ahrens, **H. Helmholz**, D. L. Alon, and S. V. Amengual Garí, “Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer”. *Journal of the Acoustical Society of America, vol. 150, no. 2*, Aug. 2021, pp. 805-815, DOI: 10.1121/10.0005754.

[L] T. Deppisch, **H. Helmholz**, J. Ahrens, “End-to-End Magnitude Least Squares Optimization for Binaural Rendering of Spherical Microphone Array Signals”. *International Conference on Immersive and 3D Audio*, Bologna, Italy, Sep. 2021, pp. 1-8.

[M] J. Ahrens, **H. Helmholz**, D. L. Alon, and S. V. Amengual Garí, “A Head-Mounted Microphone Array for Binaural Rendering”. *International Conference on Immersive and 3D Audio*, Bologna, Italy, Sep. 2021, pp. 1-7.

[N] J. Ahrens, **H. Helmholz**, D. L. Alon, and S. V. Amengual Garí, “Spherical Harmonic Decomposition of a Sound Field Based on Microphones Around the Circumference of a Human Head”. *Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, Oct. 2021, pp. 1–5.

Acknowledgments

I want to thank Jens Ahrens for his academic supervision, personal guidance and invaluable contributions to the research throughout this project. Besides, thank you to Wolfgang Kropp and others in the team at Applied Acoustics for providing a supportive environment in dire times.

Furthermore, I want to acknowledge my colleagues at Applied Acoustics, Laura, Nata, Nara, Georgios, Georgios, Carmen, Fati, Leon, as well as the other fellow PhD students out there, for sharing this experience. Special thanks to Carl Andersson, Tim Lübeck and Thomas Deppisch for their helpful insights and discussions on this research.

I want to highlight the dedication of my colleagues at the *Junge DEGA* and the *EAA Young Acousticians Network*, engaging young people with academia and acoustics at so many events over the years.

The work in this thesis has been accomplished from May 2018 to October 2021 at the Division of Applied Acoustics at Chalmers University of Technology within the research project *"Preservation of Perceived Spaciousness in Binaural Rendering of Spherical Microphone Array Recordings"*. Thank you to Ravish Mehra, Philip Robinson and the rest of the team at *Facebook Reality Labs Research* (former *Oculus Research*) for endorsing and financing this project. Special thanks to Sebastià V. Amengual Garí and David Lou Alon for consistently supporting and promoting our work.

Thank you to my friends and family for the continued support throughout this distant period in life. In particular, I want to thank Noemi for not giving up on herself or myself and for having shared the better parts of this journey.

Finally, I want to thank the COVID-19 pandemic for nothing.

Contents

Part I

Overview

CHAPTER 1

Introduction

This chapter provides a concise explanation of the fundamentals of using spherical microphone arrays for the binaural reproduction of spatial sound fields. The goal is to introduce the relevant terminology and concepts to non-experts of the field. For a detailed treatment of the mathematical concepts and discussion of the challenges at hand, the reader is referred to the referenced scientific literature including contributions by the author of this thesis.

1.1 Context

In recent years, a growing demand and the availability of hardware and software for the creation and consumption of *extended reality* (XR) content can be seen. XR applications may provide improved immersion, spatial fidelity as well as possibilities for the consumer to interact with and move around within the presented environment. Thereby, more and more sophisticated technologies are utilized to create *virtual reality* (VR) or alter existing environments by means of *augmented reality* (AR). In doing so, the consumer's attention is often drawn towards the visual sense, while also major advancements in the processing of the auditory content are required.

In order to provide a plausible auditory representation while moving through or interacting in XR, the spatial distribution of the sound field needs to be considered. This requires more sophisticated methods for both the capture and the reproduction of the acoustic environment compared to traditional channel-based techniques like stereophony or surround sound. *Ambisonics* [1] is such a technology that allows for the recording and playback of spatial sound fields [2]. Although established in the academia for a long time, it has only recently been adopted in tools and standards for media production and broadcasting [3].

A *virtual acoustic environment* (VAE) may be either created, simulated or captured from artificially composed as well as physically existing scenarios. The reproduced spatial sound field is typically presented either to an audience area via an array of loudspeakers or to a single listener via headphones. The latter is termed binaural rendering, which facilitates the ability of the auditory system to localize sounds in a three-dimensional space by the means of two

Figure 1.1: *Neumann KU100* dummy head to capture HRIR data under free-field conditions in an anechoic chamber. Image from [4] licensed under `CC BY 3.0`.

ears. A comprehensive overview over different techniques to achieve binaural reproduction can be found for example in [5], [6].

In binaural hearing, humans as well as other mammals utilize auditory cues like *inter-aural time differences* (ITD) and *inter-aural level differences* (ILD) to conceive their acoustic environment [7]. Such cues may be encoded in the form of acoustic transfer paths from a sound source at different incidence directions to the listener's ear drum, which are termed *head-related impulse responses* (HRIRs) in a free-field i.e., anechoic environment. Binaural cues are thereby highly specific to the listener according to the individual shape of their outer ears (pinnae), head, shoulders and torso. This personalisation is a major challenge in the application of binaural rendering, as capturing the individual HRIRs of every listener is currently unfeasible. Academic research and commercial applications therefore often use the data of acoustic dummy heads, like the one shown in Fig. 1.1, the render non-individualized binaural sounds. Thereby, various databases exist, e.g. [1], that make generic as well as individual HRIR data sets with very high spatial resolution freely available to the public.

The use of generic (or mismatched) HRIRs over personal ones has diverse implications on the timbral and spatial accuracy of the reproduced acoustic scene [8], [9]. For example, the reproduced virtual source usually shows deviations in it's perceived direction that is larger in terms of elevation than in azimuth compared to the real-world sound source, when using non-individualized HRIRs [10]. Typical problems of a static binaural auralization i.e., front-back confusion and a lack of externalization of the rendered sound source, can be substantially mitigated when applying dynamic head-tracking [11]. Latter refers to incorporating the listener's instantaneous head orientation into the rendering process, which results in an anchored (world-locked) auralization of the sound scene. This allows the listener to move freely within the acoustic environment, although an accurate reproduction is currently limited to head rotations around up to three axis i.e., *three degrees-of-freedom* (3-DOF).

[1] `https://www.sofaconventions.org/mediawiki/index.php/Files`

1.2 Applications

This thesis focuses on the application of capturing a real-world acoustic environment by means of a *spherical microphone array* (SMA). The goal is to create a head-tracked binaural reproduction via headphones that is indistinguishable from natural hearing in the given scenario.

Under room conditions, more binaural cues than fundamental ITDs and ILDs are available to the listener to accurately localize a sound source with respect to distance and incidence direction. There, the auditory system also utilizes for example the time structure, loudness and coloration of reflections of the sound on objects within and the boundaries of the room. In analogy to anechoic HRIRs, the transfer paths from a sound source to the ears including the room's influence are then termed *binaural room impulse responses* (BRIRs).

Figure 1.2: Dummy head (audience area, back center) to capture BRIR data from a sound source (stage, front right) under room conditions. Image from [12] licensed under `CC BY 4.0`.

BRIRs accurately preserve the characteristics of both the sound source (e.g. position, timbre, directivity) and of the room (e.g. modes, reflections, reverberation). An exemplary measurement setup is shown in Fig. 1.2. This method is capable of providing an authentic i.e., perceptually indistinguishable from reality, reproduction of the the captured space for individual head orientations [13], as well as a dense grid of orientations [14] or positions [15]. This method for the binaural reproduction of a captured space is represented by the top configuration in Fig. 1.3. A single-channel signal may be convolved with the BRIRs from a sound source in order to auralize arbitrary (typically anechoic recorded) audio content in the captured environment.

Measured BRIRs directly encode the listener-specific cues of the utilized dummy head in the captured ear signals. This prohibits the use of individualized HRIRs during the reproduction with this auralization method. It is therefore desired to first capture the spatial composition of the sound field without any binaural cues. This may be achieved by acquiring the transfer paths from the sound source to a microphone array at the listening location, termed *array room impulse responses* (ARIRs). A device that allows to measure ARIRs with arbitrarily high spatial resolution by sequentially capturing the individual microphone positions is shown in Fig. 1.4. Convolution of an arbitrary source

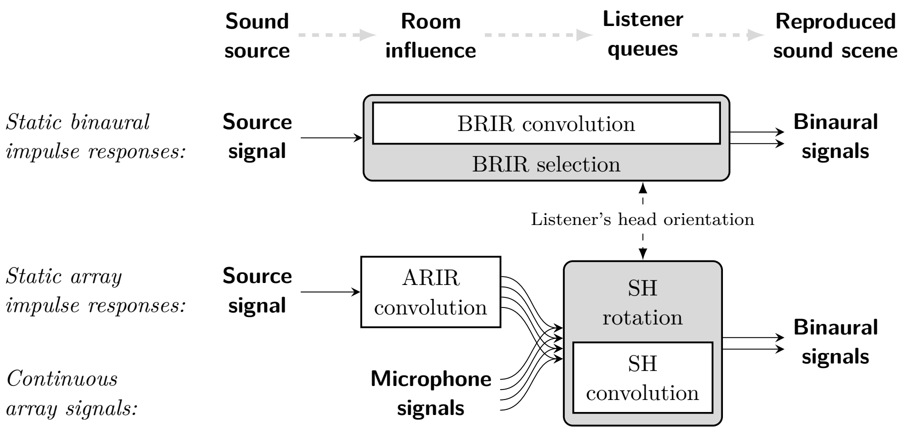

Figure 1.3: Methods for the spatial reproduction of an existing sound field by means of binaural rendering. From top to bottom, the methods provide increasing degrees of freedom during capture and reproduction.

signal with the ARIRs results in the equivalent number of microphone array signals, as shown in the middle configuration in Fig. 1.3.

For SMAs, convenient mathematical tools exist that allow for the extraction of a sound field description from the microphone signals by means of plane wave decomposition [17]. Thereby, the sound pressure distribution on the surface of the array is transformed into the *spherical harmonic* (SH) domain by means of the spatial Fourier transform [18], [19]. SHs form a set of mathematically orthogonal basis functions, with the first real SHs up to 3rd order visualized in the cover image of this thesis.

Subsequently, different methods exist to binaurally render the decomposed sound field, e.g. by an intermediate decoding to virtual loudspeakers [20]

Figure 1.4: *VariSphear* SMA measurement device to capture ARIR data [5]. Image from [12] licensed under `CC BY 4.0`.

Figure 1.5: *mh acoustics Eigenmike em32* SMA [16]. [Image: Jens Ahrens]

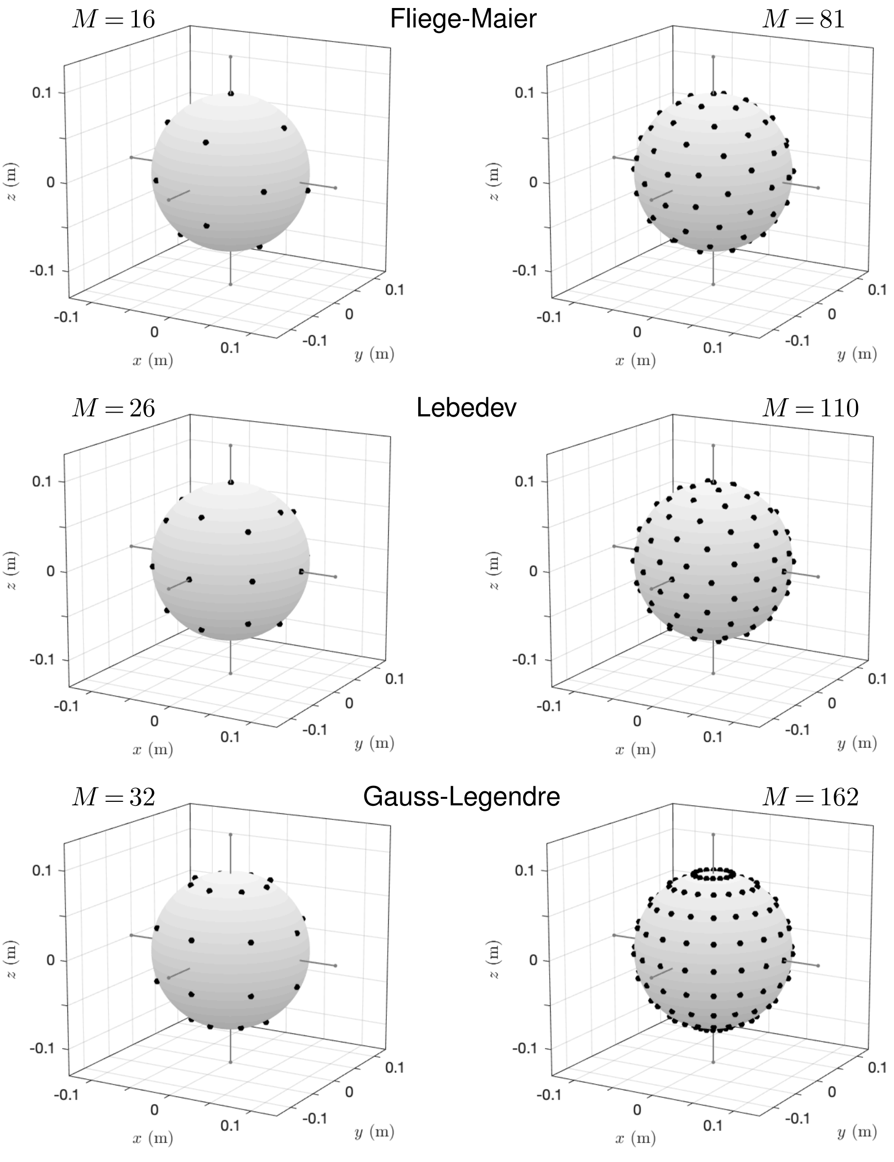

Figure 1.6: Exemplary SMA sampling grids with differing numbers of microphones M, suitable for the decomposition of the sound field at a maximum SH order of 3 (left) and 8 (right).

or by a combination with the similarly transformed HRIR data directly in the SH domain [21]. In this work, the latter method is implemented in an end-to-end real-time processing pipeline as documented in Paper [A] and Paper [B].

An example for a commercially available SMA is shown in Fig. 1.5, whose 32 microphones allow for a sound field decomposition of up to 4th SH order. A multitude of sampling strategies exist to yield a grid of microphone positions which compromise between the stability when extracting the SH coefficients and the sampling efficiency [22]. Fig. 1.6 visualizes the resulting microphone positions on spherical grids following some exemplary *Fliege-Maier* [23], *Lebedev* [24] and *Gauss-Legendre* [25] sampling strategies.

The two presented approaches based on BRIRs and ARIRs do not allow for changes in the spatial composition of the sound field, which is encoded in the captured impulse responses as a *linear time-invariant* (LTI) system. This means, apart from the head-tracked orientation of the listener, the acoustic scene remains static. In order to allow for the reproduction of numerous sources, varying environmental conditions as well as changes in the position of the sound sources or the SMA receiver, the continuous stream of microphone array signals needs to be processed. Such continuous signals may be live-captured from an SMA, replayed from local storage medium or streamed from an internet service.

At its core, the rendering of the continuous SMA signals is identical to the reproduction with ARIRs, as indicated by the bottom configuration in Fig. 1.3. In order to incorporate the listener's head orientation and to not cause artifacts or dropouts, the rendering needs to be performed in real-time i.e., in a block-wise manner. This means the processing of a block of signals needs to be completed before the arrival of the next block i.e., in quick succession of block lengths of typically tens of milliseconds. To avoid long block lengths and therefore potentially noticeable delays in an auralization, sophisticated processing techniques like partitioned convolution may need to be implemented [26].

1.3 Challenges

The mathematical and physical concepts of capturing spatial sound fields with SMAs is well understood [27]. However, real-world applications induce considerable physical limitations over how they are described in the theory. This section provides a brief outline of the existing challenges regarding the accuracy of the produced ear signals in the binaural rendering of SMA signals.

1.3.1 Spatial Undersampling

The theory requires a continuous layer of microphones whereas practical implementations may only employ a finite set of discrete microphones. This undersampling of the sound field causes spatial aliasing at high frequencies [28]. This results in spatial ambiguities which may be described as a loss in spatial resolution, e.g. an unnatural increase in perceived size or proximity of the reproduced sound source. On average, the spatial aliasing also leads to an over-emphasis of the magnitude spectrum at high frequencies, which may be compensated by a matched equalization filter [5].

Furthermore, the limited number of array sensors allows for the extraction of the captured sound field of only lower SH orders. With the employed rendering method, this results in the higher-order components of the HRIR data being discarded. Therefore, this is termed SH order truncation and may be described as a loss in spatial resolution i.e., a spatial smoothing, of the listener-specific cues. On average, the effect also causes an attenuation of the signals at high frequencies, which again may be addressed by global equalization [29].

The greatest challenge with the SH order truncation lays in deviations from the high-resolution data which are strongly dependent on the direction of sound incidence. To visualize this, Fig. 1.7 and Fig. 1.8 show the resulting errors varying considerably over frequency and head orientation for different SMA configurations according to Fig. 1.6 at 3rd and 8th SH order. The individual spectra were smoothed in third octave bands before the differences to the utilized *Neumann KU100* [4] dummy head reference are determined.

The depicted examples render a simulated plane wave under anechoic conditions impinging from frontal direction. The resulting deviations are qualitatively similar from measurements of SMAs under real-world room conditions (e.g. [37, Fig. 6]). A higher SH order raises the aliasing frequency as apparent when comparing Fig. 1.7 and Fig. 1.8. The several spherical sampling grids

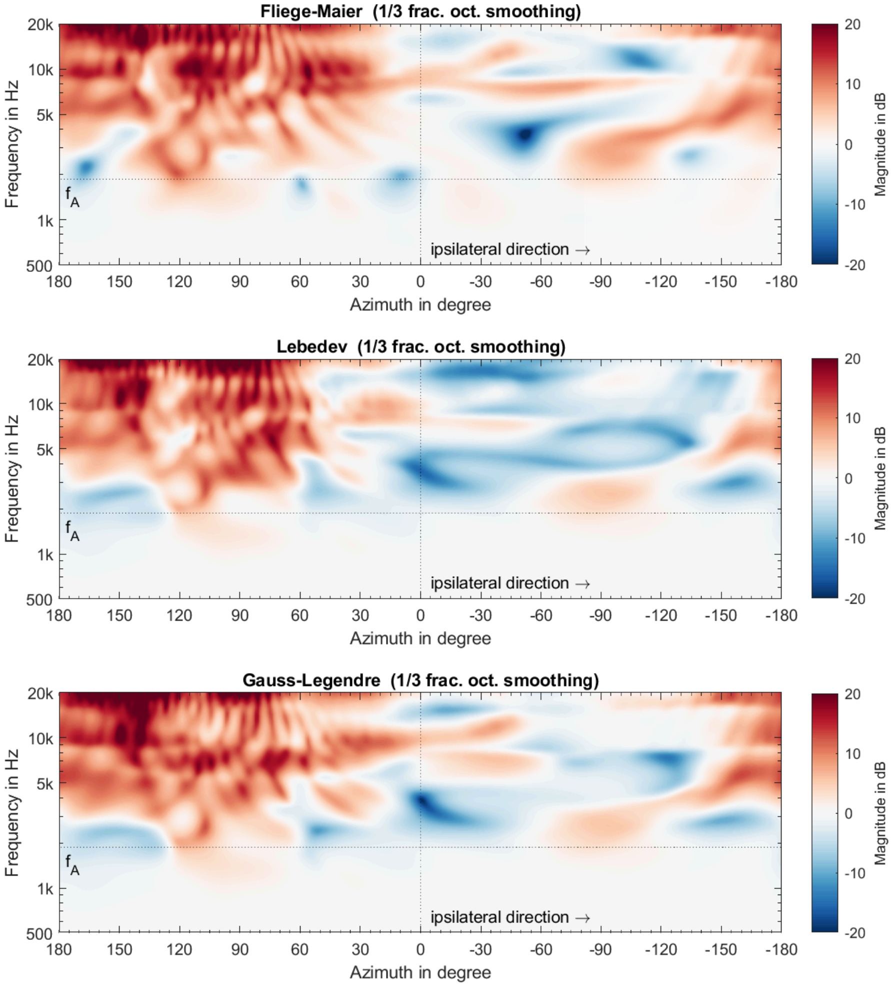

Figure 1.7: Deviations for the left ear from a simulated plane wave from azimuth 0° in the horizontal plane on SMAs with different sampling grids at 3rd SH order compared to the corresponding dummy head reference at 44th SH order. The strong dependency on the azimuth angle of head orientation is apparent above the respective aliasing frequency $f_{\mathrm{A}} = 1.9\,\mathrm{kHz}$.

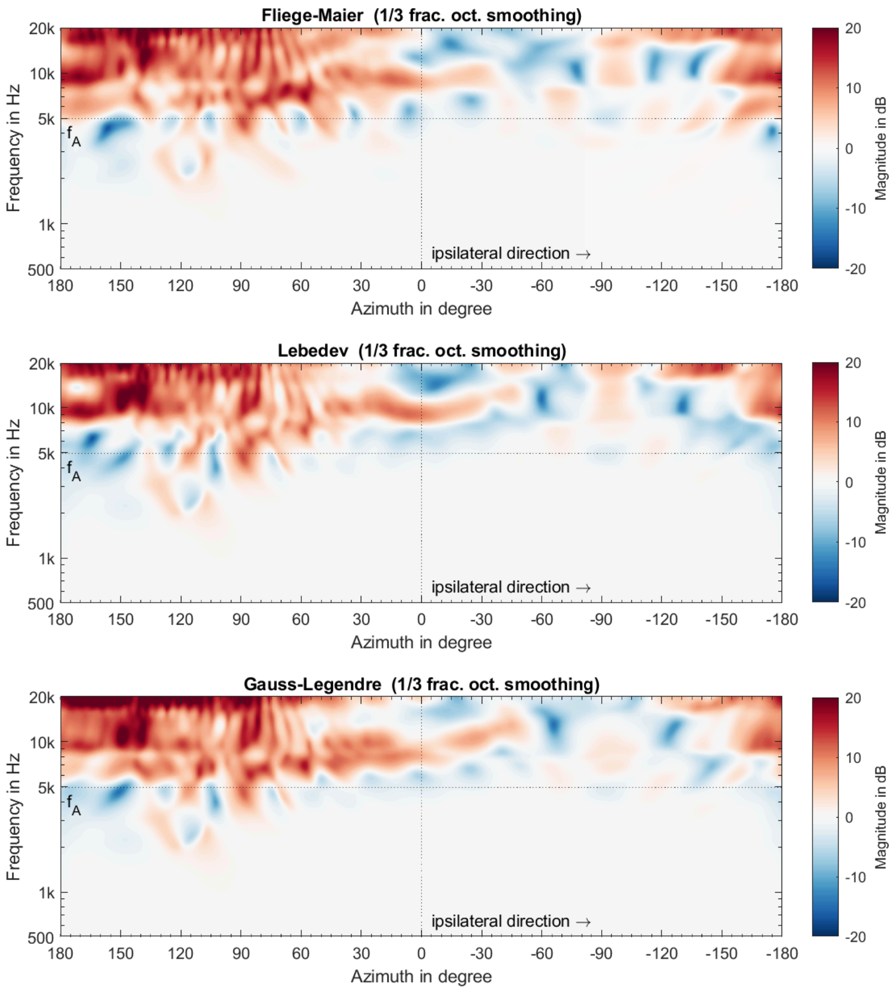

Figure 1.8: Deviations for the left ear from a simulated plane wave from azimuth 0° in the horizontal plane on SMAs with different sampling grids at 8th SH order compared to the corresponding dummy head reference at 44th SH order. The strong dependency on the azimuth angle of head orientation is apparent above the respective aliasing frequency $f_{\mathrm{A}} = 5\,\mathrm{kHz}$.

show slight differences in detail but yield qualitatively similar results. Also other parameters like the employed radial filters (cf. Sec. 1.3.2) may exert a minor influence on the accuracy of the rendered target sound field ear signals.

The local deviations of the order truncated reproduction are caused by spatial undersampling of the naturally very complex directivity of the HRIR data. Thereby, the largest errors arise at high frequencies and for contralateral sound incidence i.e., in the ear facing away from the source (cf. positive azimuth angles in Fig. 1.7 and Fig. 1.8). These local deviations cannot be mitigated by means of a global equalization of the magnitude spectra, but require more sophisticated processing methods. For this, tapering of the SH coefficients [30] as well as time alignment [31] and *Magnitude Least-Squares* (MagLS) optimization [32] of the HRIR data have been proposed. An improved version of the MagLS approach, which also considers the spatial aliasing on the SMA scattering body as well as non-optimal radial filters (cf. Sec. 1.3.2) into the optimization, has been recently proposed in Paper [L, 33].

The perceived fidelity in timbral and spatial reproduction of different SMA and rendering configurations has been investigated, e.g. in [5], [34] as well as Paper [F, 35]. The perceived effectiveness of different mitigation approaches for spatial aliasing and SH order truncation has been compared, e.g. with contributions of this author in Paper [G, 36], Paper [H, 37] and Paper [I, 38].

1.3.2 Array Imperfections and Additive Noise

Hardware imperfections may constitute substantial limitations on the theoretical models for the processing of SMA signals. On the one hand, imperfect placement or a sensitivity mismatch of the microphones may cause spatial distortions of the captured sound field, since the acquired information does not coincide with the desired spatial sampling. Considering practical constrains on the accuracy of the microphone positioning showed a relevant influence, which should be considered in the design and manufacturing of SMAs [28].

On the other hand, additive noise may present a considerable nuisance factor in practical applications of SMAs. The theoretical influence of sensor self-noise was investigated in simulations of a spatially uniform *white-noise-gain* (WNG), e.g. in [5], [28], [39]. This work is continued in Paper [C], Paper [D] and Paper [E] for a practical implementation of the SMA processing pipeline as well as for spatially non-uniform noise contributions from the individual array sensors. The practical relevance of various noise contributions in different real-world applications is briefly discussed in Sec. 3.1.

The processing in the SH domain allows for removing the influence of the scattering of the microphone array body from the captured signals. This is achieved by the so-called radial filters, which are part of the *"SH convolution"* process in Fig. 1.3. Open arrays may be realized, whereas the SMA designs with a rigid scattering body (e.g. Fig. 1.4 and Fig. 1.5) have been preferred for their improved numerical conditioning of the resulting filters [28].

In theory, the radial filters require high gains at low frequencies and therefore amplify the influence of array imperfections and additive noise. Various approaches have been proposed to limit the radial filter gains and equalize the magnitude transfer function, e.g. by Tikhonov-regularization [40], highpass filtering [41] and soft-clipping [5]. Latter method is applied in this work and is visualized aside the unlimited filters in Fig. 1.9. The limitation may be interpreted as a reduction of the available SH order of the sound field at low frequencies (cf. Sec. 1.3.1). This is equivalent to a loss in spatial resolution, which is perpetually irrelevant for reasonably chosen gain limits due to the low directivity of HRIR data at low frequencies [31]. Furthermore, the impact of a conservative limitation on the magnitude spectra of the binaural output signals at low and high frequencies may be compensated by global equalization [37].

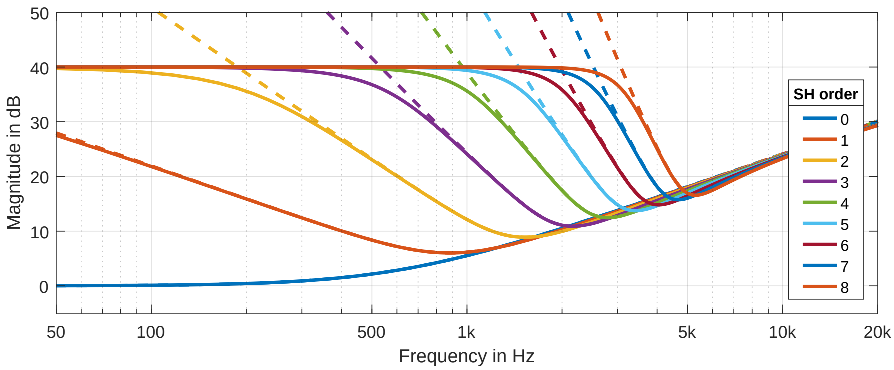

Figure 1.9: Exemplary radial filters applied for the decomposition of the sound field from an SMA with a rigid spherical scatterer of radius 8.75 cm. The very high amplification gains at low frequencies (dashed lines) require regularization in practical applications (solid lines). The plot visualizes magnitude soft-clipping [5] to 40 dB.

1.4 Objectives

This thesis is focused on the following key objectives:

- Implementation of a Python application for the binaural rendering of streamed audio signals from spherical microphone arrays.
  → Paper [A], Paper [B]

- Identification of the array and rendering requirements for a high reproduction quality with respect to perceived spaciousness and timbre.
  → Paper [F, 35], Paper [G, 36], Paper [H, 37], Paper [I, 38]

- Assessment of additive noise (e.g. microphone self-noise) in the array signals and its influence on the rendered ear signals.
  → Paper [C], Paper [D], Paper [E]

1.5 Thesis Outline

This thesis is structured as follows:

Chapter 1 presented a brief introduction of the terminology and concepts employed for the binaural rendering of spherical microphone arrays signals.

Chapter 2 provides an individual summary of the papers included in Part II.

Chapter 3 concludes this thesis and gives a perspective on future work.

CHAPTER 2

Summary of Included Papers

[A] **Hannes Helmholz**, Carl Andersson, Jens Ahrens
Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals
Published in Fortschritte der Akustik – DAGA 2019,
pp. 1462-1465, Rostock, Germany, Mar. 2019.
©2019 Deutsche Gesellschaft für Akustik .

This publication presents the first results of our implementation for the binaural rendering of signals from spherical microphone arrays. The software is based on previously published Matlab and Python toolboxes which were also used to verify the results. The publication shows that Python, together with the utilized programming techniques and libraries, can provide a suitable framework for such processing-heavy audio applications, even under real-time constraints. The implementation at the time is able to render array signals of tens to hundreds of microphone channels on conventional computer hardware. The work will be refined in Paper [B] and provides the foundation for the following publications. The implementation was made publicly available under an open-source license.

[B] **Hannes Helmholz**, Tim Lübeck, Jens Ahrens, Sebastià V. Amengual Garí, David Lou Alon, Ravish Mehra
Updates on the Real-Time Spherical Array Renderer (ReTiSAR)
Published in Fortschritte der Akustik – DAGA 2020,
pp. 1169–1172, Hannover, Germany, Mar. 2020.

This publication presents improvements to our in Paper [A] published binaural renderer for spherical microphone array signals. These comprise for example an interface for impulse response data sets in the SOFA format as well as the methods for the mitigation of processing limitations due to spatial undersampling. Furthermore, we implemented the addition of noise into the microphone channels of any kind of spherical array configuration with arbitrary coloration and inter-channel distribution. This functionality will be used in the following Papers [C], [D], [E] for the detailed assessment of additive noise components. By now, performance improvements to the processing pipeline allow for the real-time rendering of hundreds of microphone impulse responses with the simultaneous emulation of additive noise to all channels.

[C] **Hannes Helmholz**, Jens Ahrens, David Lou Alon, Sebastià V. Amengual Garí, Ravish Mehra
Evaluation of Sensor Self-Noise In Binaural Rendering of Spherical Microphone Array Signals
Published in Proceedings of International Conference on Acoustics, Speech and Signal Processing,
pp. 161-165, Barcelona, Spain, May 2020.

This work presents the first results of auralizing and analyzing the propagation of additive noise through a real-time processing pipeline for spherical microphone arrays, as published previously in Paper [B]. The instrumental evaluation confirms that different array and rendering parameters have a strong influence on the spectral balance and the overall level of the rendered noise. The character of the noise is thereby independent of the listener's head orientation in the case of spatially uniformly distributed noise. However, the timbre changes with head orientation in the case of spatially non-uniform noise contributions, e.g. from a single channel exhibiting an increased noise level. In a perceptual user study presenting such a scenario, we determine audibility thresholds of the coloration artifact during head rotations, which vary considerably for different array configurations. The perceptual thresholds will be utilized to develop a predictive model for audibility in Paper [E].

[D] **Hannes Helmholz**, David Lou Alon, Sebastià V. Amengual Garí, Jens Ahrens
Instrumental Evaluation of Sensor Self-Noise in Binaural Rendering of Spherical Microphone Array Signals
Published in Proceedings of Forum Acusticum,
pp. 1349–1356, Lyon, France, Dec. 2020.
©2020 European Acoustics Association DOI: 10.48465/fa.2020.0074 .

We pursue the methodology of Paper [C] to instrumentally investigate noise components in the binaural rendering of spherical microphone array signals. Here, the focus is on examining the frequency-dependent effects on the wanted target signals versus the unwanted noise signals. The analysis is performed by comparing the respective *signal-to-noise ratio* (SNR) between a selected microphone of the array and the binaural output signals. We find that in particular small array radii or a weak restriction of the radial filter gains can impair the SNR in the rendered signals. Array configurations larger in size or number of sensors on the other hand can yield an improved SNR compared to a single sensor. The observed effects are strongly frequency-dependent while influencing the SNR by up to $\pm 10\,\mathrm{dB}$. We also find some undesired variations in the spectral properties of rendered additive noise during head rotations for some higher-order *Lebedev* grids. In the appendix, we measure and analyse the self-noise signals from a commercially available spherical microphone array, which will be utilized in Paper [E]. The results show that the observed noise levels, color and inter-channel differences strongly depend on the specific hardware as well as the employed pre-amplification gain.

[E] **Hannes Helmholz**, David Lou Alon, Sebastià V. Amengual Garí, Jens Ahrens
Effects of Additive Noise in Binaural Rendering of Spherical Microphone Array Signals
Accepted for publication in
IEEE/ACM Transactions on Audio, Speech, and Language Processing, Oct. 2021.

We conclude the investigation of additive noise components in streamed signals from spherical microphone arrays. This publication is focused on generalizing our previous findings to a broader set of array configurations and noise distributions. wider range of array and rendering configurations. We calibrate the instrumental metric of *Composite Loudness Level* (CLL) to the perceptual thresholds determined in Paper [C] and predict the audibility of changes in the rendered noise due to head rotations. In this way, we demonstrate that in particular spherical sampling grids that exhibit negative quadrature weights can produce audible variations even if the noise level is equal in all channels. This is seen for example in some *Lebedev* and *Fliege-Maier* grids, as observed before in Paper [D]. Overall, the analysis shows that the influence of the noise from individual array channels is determined by the proximity of their virtual location to the relative trajectory of the ears.

CHAPTER 3

Concluding Remarks and Future Work

This chapter provides a conclusion on the objectives presented in Sec. 1.4.

An implementation of a processing pipeline for the binaural rendering of streamed audio signals from spherical microphone arrays was presented. The Python programming language was found to be a viable framework for real-time audio processing applications, even when hundreds of microphone channels have to be processed simultaneously (cf. Paper [A]). Optimization of the implementation allow for the reproduction of higher-order SMA configurations or otherwise at a lower rendering block-size with the latter providing a lower head-tracking and overall processing latency (cf. Paper [B]). The utilized frame size in the block-wise processing does not have an audible effect of the expansion of the microphone array signals into SH coefficients or other components of the rendering pipeline.

SMA rendering configurations have been perceptually compared to the corresponding auralizations of a dummy head reference. In direct comparison, differences in terms of spaciousness and / or timbre remain audible up to 12th SH order (cf. Paper [F, 35], [34]). Multiple algorithms to mitigate the perceptual impairments from spatial undersampling, e.g. *Spherical Head Filter* (SHF) [29] and MagLS [32], yielded comparable but significant improvements at low to medium SH orders in real-world listening environments (cf. Paper [G, 36], Paper [H, 37], Paper [I, 38]). Thereby, the SH order truncation seems to wield a larger influence than the apparent spatial aliasing in the binaural rendering of SMA data.

3.1 Relevance of Noise in Practical Applications

The implemented processing pipeline was extended to generate noise of arbitrary coloration, which may be added to the microphone channels of any SMA configuration. The presented work provides tools to determine the influence of noise in diverse spatial contributions on the rendered ear signals. Table 3.1 gives an overview of potential signal components in the SMA signals in a practical recording application. All different types of noise are likely to be present in a real-world scenario, while their relative strengths highly depend on the environmental conditions as well as the employed recording hardware.

The wanted components of the SMA signals are samples of the impinging sound field, which are related through the wave equation. Thereby, acoustic environmental noise is part of the captured target source signals, with both components adding up coherently between sensors in the rendering according to the SMA literature (cf. Sec. 1.3.2). Nevertheless, the two sound field components may be modeled assuming differing long-term average spectra (cf. coloration in Table 3.1) and representative signal levels (e.g. [42]–[44]).

The captured sound field will be superimposed with varying amounts of additive noise, which are mostly assumed to be uncorrelated. This causes them to add up incoherently in the binaural rendering, resulting in an alteration of the apparent *signal-to-noise ratio* (SNR) depending on the hardware and software configuration (cf. Paper [D]). Additive noise may originate from pre-amplifiers, *analog-to-digital converters* (ADCs), signal quantization, thermal noise in the microphone capsules or electromagnetic interference like hum. They may be modeled depending on their spatial distribution as *uniform*, *normal* or *non-uniform* contributions to the SMA signals (cf. Paper [C] and Paper [E]).

Any undesired variations in the rendered additive noise over different head orientations will be masked by the present sound field components. Audibility of the variations may therefore be very limited in the presence of acoustic background noise. A conclusive assessment of the practical relevance in representative recording applications is subject to future work.

	Sound field		**Additive noise**		
	Target source signals	**Environ-mental noise**	**Input chain** pre-amp + ADC + quantization	**Microphone capsule**	**Electro-magnetic interference**
Average coloration	pink to red [45], [46]	white to red [43], [47]	white + pink (flicker) [48]	white to pink → Paper [D]	tonal
Assumed correlation	correlated		uncorrelated		correlated
Contribution to binaural signals	inter-channel differences due to wave propagation → SMA literature → **SNR:** Paper [D]		**Uniform:** resulting in potentially non-unif. distribution depending on the sampling grid → Paper [E, Sec. 4] **Normal:** under consideration of capsule sensitivity equalization → Paper [E, Sec. 5]		diverse e.g. **Non-uniform** → Paper [E, Sec. 3]

Table 3.1: Categorization of different target and noise components present in the SMA signals.

3.2 Simplification of the Array Geometry

Binaural rendering of a sound field captured by an SMA has proven to be a convenient framework. However, the amount of microphones for a high-fidelity reproduction can be substantially high. Even with state-of-the-art rendering techniques, SH orders of 8 or higher may be required to render a captured sound field with negligible perceptual impairments [5], [31], [34], [37], [38]. This necessitates an SMA with at least 81 microphones (cf. Paper [E, Table 2]), which renders such a design currently unviable to be built for commercial applications. It is obvious that there is a desire to lower the number of microphones while maintaining a similar reproduction quality. For the discussed application of physically accurate binaural rendering in SHs, this may be achieved by:

- Mitigate the impairments of SH truncation of the HRIR data, which may allow for a significant reduction of the required SH order and therefore the number of sensors in the deployed SMA. A method was proposed for example in [49], [50], although it currently does not allow for similar a freedom in terms of reproducing with dynamic head-tracking of the listener and/or capturing dynamic scenes from streamed microphone signals (cf. Fig. 1.3).

- Extract the high-resolution description of the sound field from an array geometry employing less microphones. This was attempted many times, e.g. for planar [51], [52], cylindrical [53], [54] and spherical [55], [56] microphone arrays, but did not yield a SH-compatible representation of the three-dimensional sound field.

Capturing the sound field with an SMA has the benefit of providing equal properties for all angles of sound incidence. However, there may be scenarios where this might not be required, as the human auditory system is less sensitive for spatial differences in elevated sound sources [7], [58]. Therefore, it may be sufficient to focus on an accurate reproduction of sources in the horizontal plane [54], [59].

A solution based on and microphones located solely on the equator of a spherical scattering body, termed *equatorial microphone array* (EMA), was proposed with contributions of this author in Paper [J, 60] and Paper [K, 61]. This was adapted to scattering geometries in the shape of a human head in Paper [M, 57] and Paper [N, 62], termed XMA. A visualization of possible

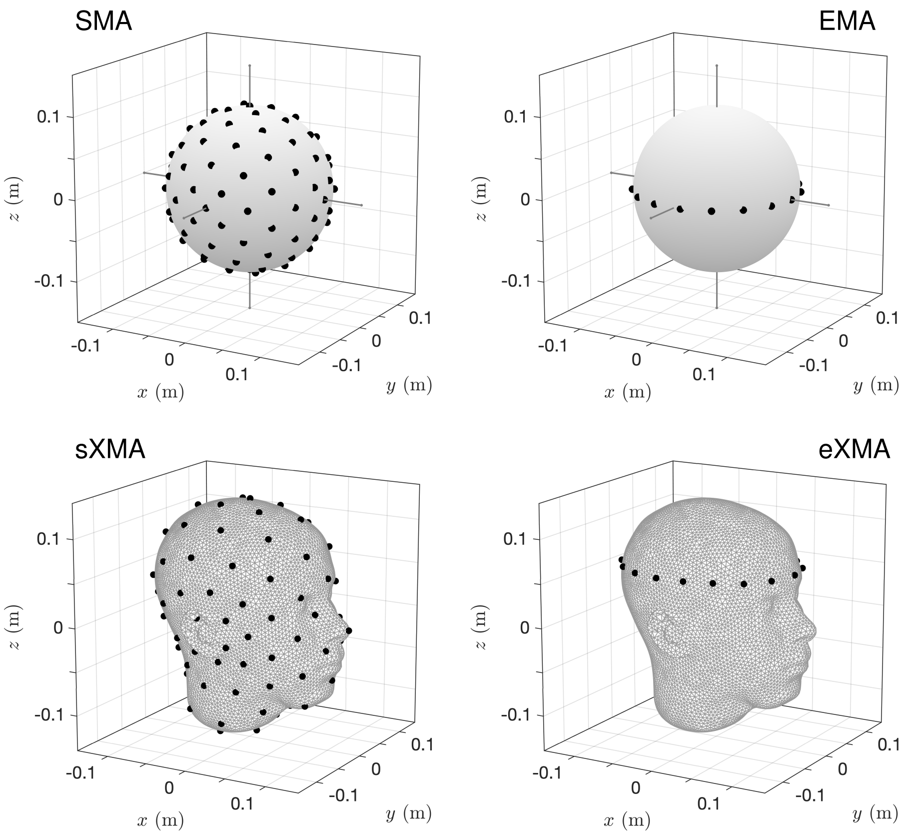

Figure 3.1: Exemplary microphone arrays for 8th-order decomposition of the sound field. **SMA:** 110 microphone Lebedev grid (top left), **EMA:** 17 equiangularly spaced microphones (top right), **sXMA:** 110 microphone Lebedev-inspired grid (bottom left), **eXMA:** 18 equiangularly spaced microphones (bottom right). Plots reprinted from [57, Fig. 1 and Fig. 2].

microphone array configurations for the different array geometries is given in Fig. 3.1.

The evaluation of the proposed array geometries was performed based on simulations of anechoic plane waves and point sources. In these scenarios, the different approaches yield very promising results with only mild deviations from the optimal results, even for sound sources outside of the horizontal plane. A perceptual evaluation of the different array geometries based on measurements of real-world acoustic scenarios will be performed in the future. The study will investigate how the spatial fidelity of the room reproduction, in particular from elevated sound sources, is affected by the different array geometries (SMA, EMA, XMA in comparison to reference binaural cues).

The work on equatorial microphone arrangements as well as potentially head-mounted microphone arrays will be presented and discussed in the PhD dissertation of the author of this thesis.

References

[1] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973, ISSN: 00047554.

[2] D. Malham, "Higher Order Ambisonic Systems for the Spatialisation of Sound," in *International Computer Music Conference Proceedings*, Beijing, China: Michigan Publishing, 1999, pp. 484–487.

[3] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording*. Cham, Switzerland: Springer Open, 2019, pp. 1–210, ISBN: 9783540786115. DOI: 10.1007/978-3-030-17207-7.

[4] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *Fortschritte der Akustik – AIA/DAGA 2013*, Meran, Italy: Deutsche Gesellschaft für Akustik, 2013, pp. 592–595. DOI: 10.5281/zenodo.3928296.

[5] B. Bernschütz, "Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording," PhD Thesis, Technische Universität Berlin, Germany, 2016, pp. 1–284. DOI: 10.14279/depositonce-5082.

[6] M. Otani, H. Shigetani, M. Mitsuishi, and R. Matsuda, "Binaural Ambisonics: Its Optimization and Applications for Auralization," *Acoustical Science and Technology*, vol. 41, no. 1, pp. 142–150, 2020, ISSN: 1346-3969. DOI: 10.1250/ast.41.142.

[7] J. Blauert and W. Lindemann, "Spatial Mapping of Intracranial Auditory Events for Various Degrees of Interaural Coherence," *Journal of the Acoustical Society of America*, vol. 79, no. 3, pp. 806–813, 1986. DOI: 10.1121/1.393471.

[8] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993, ISSN: NA. DOI: 10.1121/1.407089.

[9] F. Zagala, M. Noisternig, and B. F. Katz, "Comparison of Direct and Indirect Perceptual Head-Related Transfer Function Selection Methods," *Journal of the Acoustical Society of America*, vol. 147, no. 5, pp. 3376–3389, 2020, ISSN: 0001-4966. DOI: 10.1121/10.0001183.

[10] Z. Ben-Hur, D. L. Alon, P. W. Robinson, and R. Mehra, "Localization of Virtual Sounds in Dynamic Listening Using Sparse HRTFs," in *AES International Conference onAudio for Virtual and Augmented Reality*, Online: Audio Engineering Society, 2020, pp. 1–9.

[11] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. Katz, and C. de Boishéraud, "Influence of Head Tracking on the Externalization of Speech Stimuli for Non-Individualized Binaural Synthesis," *Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 2011–2023, 2017, ISSN: 0001-4966. DOI: 10.1121/1.4978612.

[12] P. Stade, B. Bernschütz, and M. Rühl, "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," in *27th Tonmeistertagung – VDT International Convention*, Cologne, Germany: Verband Deutscher Tonmeister e.V., 2012, pp. 551–567.

[13] J. Oberem, B. Masiero, and J. Fels, "Experiments on Authenticity and Plausibility of Binaural Reproduction via Headphones Employing Different Recording Methods," *Applied Acoustics*, vol. 114, pp. 71–78, 2016, ISSN: 1872910X. DOI: 10.1016/j.apacoust.2016.07.009.

[14] F. Brinkmann, A. Lindau, and S. Weinzierl, "On the Authenticity of Individual Dynamic Binaural Synthesis," *Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 1784–1795, 2017, ISSN: 0001-4966. DOI: 10.1121/1.5005606.

[15] A. Neidhardt and B. Reif, “Minimum BRIR Grid Resolution for Interactive Position Changes in Dynamic Binaural Synthesis,” in *AES Convention 148*, Vienna, Austria: AES, 2020, pp. 1–10.

[16] J. Meyer and G. Elko, “A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield,” in *International Conference on Acoustics, Speech and Signal Processing*, vol. 2, Orlando, USA: IEEE, 2002, pp. 1781–1784, ISBN: 0780374029. DOI: `10.1109/ICASSP.2002.5744968`.

[17] M. Park and B. Rafaely, “Sound-Field Analysis by Plane-Wave Decomposition Using Spherical Microphone Array,” *Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3094–3103, 2005, ISSN: 0001-4966. DOI: `10.1121/1.2063108`.

[18] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, 1st ed., E. G. Williams, Ed. London, UK: Academic Press, 1999, pp. 1–306, ISBN: 978-0-12-753960-7. DOI: `10.1016/B978-012753960-7/50001-2`.

[19] N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions.* Amsterdam, NL: Elsevier Science, 2005, p. 520, ISBN: 978-0080443713. DOI: `10.1016/B978-0-08-044371-3.X5000-5`.

[20] J.-M. Jot, S. Wardle, and V. Larcher, “Approaches to Binaural Synthesis,” in *AES Convention 105*, vol. 861, San Francisco, USA: Audio Engineering Society, 1998, pp. 4861–4875.

[21] B. Rafaely and A. Avni, “Interaural Cross Correlation in a Sound Field Represented by Spherical Harmonics,” *Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 823–828, 2010, ISSN: 0001-4966. DOI: `10.1121/1.3278605`.

[22] F. Zotter, “Sampling Strategies for Acoustic Holography/Holophony on the Sphere,” in *International Conference on Acoustics (NAG/DAGA)*, Rotterdam, NL: Deutsche Gesellschaft für Akustik, 2009, pp. 1107–1110.

[23] J. Fliege and U. Maier, “The Distribution of Points on the Sphere and Corresponding Cubature Formulae,” *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999, ISSN: 02724979. DOI: `10.1093/imanum/19.2.317`.

[24] V. I. Lebedev, "Spherical Quadrature Formulas Exact to Orders 25-29," *Siberian Mathematical Journal*, vol. 18, no. 1, pp. 99–107, 1977, ISSN: 1573-9260. DOI: 10.1007/BF00966954.

[25] I. Bogaert, "Iteration-Free Computation of Gauss-Legendre Quadrature Nodes and Weights," *SIAM Journal on Scientific Computing*, vol. 36, no. 3, pp. 1008–1026, 2014, ISSN: 10957200. DOI: 10.1137/140954969.

[26] F. Wefers, "Partitioned Convolution Algorithms for Real-Time Auralization," Dr. rer. nat. RWTH Aachen University, 2014, pp. 1–258, ISBN: 978-3-8325-3943-6.

[27] B. Rafaely, *Fundamentals of Spherical Array Processing*, 2nd ed., J. Benesty and W. Kellermann, Eds. Springer Berlin Heidelberg, 2015, p. 196, ISBN: 9783662456637. DOI: 10.1007/978-3-319-99561-8.

[28] B. Rafaely, "Analysis and Design of Spherical Microphone Arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005, ISSN: 10636676. DOI: 10.1109/TSA.2004.839244.

[29] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral Equalization in Binaural Signals Represented by Order-Truncated Spherical Harmonics," *Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4087–4096, 2017, ISSN: 0001-4966. DOI: 10.1121/1.4983652.

[30] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. J. Tashev, "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation," in *International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK: IEEE, 2019, pp. 261–265, ISBN: 9781538646588. DOI: 10.1109/ICASSP.2019.8683751.

[31] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural Rendering of Ambisonic Signals by Head-Related Impulse Response Time Alignment and a Diffuseness Constraint," *Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018, ISSN: 0001-4966. DOI: 10.1121/1.5040489.

[32] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Fortschritte der Akustik – DAGA 2018*, Munich, Germany: Deutsche Gesellschaft für Akustik, 2018, pp. 339–342.

[33] T. Deppisch, H. Helmholz, and J. Ahrens, "End-to-End Magnitude Least Squares Optimization for Binaural Rendering of Spherical Microphone Array Signals," in *International Conference on Immersive and 3D Audio*, Bologna, Italy, 2021, pp. 1–8.

[34] J. Ahrens and C. Andersson, "Perceptual Evaluation of Headphone Auralization of Rooms Captured with Spherical Microphone Arrays with Respect to Spaciousness and Timbre," *Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2783–2794, 2019. DOI: 10.1121/1.5096164.

[35] J. Ahrens, H. Helmholz, and C. Andersson, "Authentic Auralization of Acoustic Spaces based on Spherical Microphone Array Recordings," in *Auditorium Acoustics*, vol. 40, Hamburg, Germany: Institute of Acoustics, 2018, pp. 1–8.

[36] T. Lübeck, J. M. Arend, H. Helmholz, J. Ahrens, and C. Pörschmann, "Comparison of Mitigation Approaches of Spatial Undersampling Artifacts in Spherical Microphone Array Data Auralizations," in *Fortschritte der Akustik – DAGA 2020*, Hannover, Germany: Deutsche Gesellschaft für Akustik, 2020, pp. 623–627.

[37] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *Journal of the Audio Engineering Society*, vol. 68, no. 6, pp. 428–440, 2020. DOI: 10.17743/jaes.2020.0038.

[38] T. Lübeck, J. M. Arend, C. Pörschmann, H. Helmholz, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments Due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data: Dry Acoustic Environments," in *23rd International Conference on Digital Audio Effects*, Vienna, Austria, 2020, pp. 250–257.

[39] J. Nowak, "Quality Assessment of Spherical Microphone Array Auralizations," PhD Thesis, Technische Universität Ilmenau, Germany, 2019, pp. 1–215.

[40] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of Spherical Microphone," in *AES Convention 120*, Paris, France: Audio Engineering Society, 2006, pp. 1–24.

[41] S. Lösler and F. Zotter, "Comprehensive Radial Filter Design for Practical Higher-Order Ambisonic Recording," in *Fortschritte der Akustik – DAGA 2015*, Nürnberg, Germany: Deutsche Gesellschaft für Akustik, 2015, pp. 452–455.

[42] ITU-R BS.1116-1, *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, 1997.

[43] P. Marie, C. H. Jeong, J. Brunskog, and C. M. Petersen, "Audience Noise in Concert Halls During Musical Performances," in *41st International Congress and Exposition on Noise Control Engineering (INTER-NOISE)*, New York, USA: ICA, 2012, pp. 2231–2240, ISBN: 9781627485609. DOI: 10.1121/1.3689558.

[44] ANSI/ASA S12.2-2019, *Criteria for Evaluating Room Noise*, 2019.

[45] L. E. Cornelisse, J.-P. Gagné, and R. C. Seewald, "Long-term Average Speech Spectrum at the Chest-level Microphone Location," *Journal of Speech-Language Pathology and Audiology*, vol. 15, no. 3, pp. 7–12, 1991, ISSN: 08481970.

[46] A. Elowsson and A. Friberg, "Long-Term Average Spectrum in Popular Music and its Relation to the Level of the Percussion," in *AES Convention 142*, Berlin, Germany: Audio Engineering Society, 2017, pp. 1–12.

[47] M. Hodgson, R. Rempel, and S. Kennedy, "Measurement and Prediction of Typical Speech and Background-Noise Levels in University Classrooms During Lectures," *Journal of the Acoustical Society of America*, vol. 105, no. 1, pp. 226–233, 1999, ISSN: 0001-4966. DOI: 10.1121/1.424600.

[48] A. C. Cárdenas-Olaya, E. Rubiola, J. M. Friedt, P. Y. Bourgeois, M. Ortolano, S. Micalizio, and C. E. Calosso, "Noise Characterization of Analog to Digital Converters for Amplitude and Phase Noise Measurements," *Review of Scientific Instruments*, vol. 88, no. 6, pp. 1–10, 2017, ISSN: 10897623. DOI: 10.1063/1.4984948.

[49] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 12, pp. 2249–2262, 2019, ISSN: 23299304. DOI: 10.1109/TASLP.2019.2945479.

[50] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 29, pp. 901–913, 2021, ISSN: 23299304. DOI: 10.1109/TASLP.2021.3055038.

[51] M. Zaunschirm and F. Zotter, "Measurement Based Modal Beamforming Using Planar Circular Microphone Arrays," in *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany: EAA, 2014.

[52] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and Design of Compact Hybrid Microphone Arrays on Two-Dimensional Planes for Three-Dimensional Soundfield Analysis," *Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015, ISSN: 0001-4966. DOI: 10.1121/1.4934953.

[53] T. Betlehem and M. A. Poletti, "Measuring the Spherical-Harmonic Representation of a Sound Field Using a Cylindrical Array," in *International Conference on Acoustics, Speech and Signal Processing*, New Paltz, USA: IEEE, 2019, pp. 955–959, ISBN: 9781479981311. DOI: 10.1109/ICASSP.2019.8683701.

[54] M. B. Galindo, P. Coleman, and P. J. B. Jackson, "Microphone Array Geometries for Horizontal Spatial Audio Object Capture With Beamforming," *Journal of the Audio Engineering Society*, vol. 68, no. 5, pp. 324–337, 2020, ISSN: 15494950. DOI: 10.17743/jaes.2020.0025.

[55] T. Weller, S. Favrot, and J. M. Buchholz, "Application of a Circular 2D Hard-Sphere Microphone Array for Higher-Order Ambisonics Auralization," in *Forum Acusticum*, Aalborg, Denmark: European Acoustics Association, 2011, pp. 2269–2274.

[56] S. Koyama, K. Furuya, K. Wakayama, S. Shimauchi, and H. Saruwatari, "Analytical Approach to Transforming Filter Design for Sound Field Recording and Reproduction Using Circular Arrays with a Spherical Baffle," *Journal of the Acoustical Society of America*, vol. 139, no. 3, pp. 1024–1036, 2016, ISSN: 0001-4966. DOI: 10.1121/1.4942590.

[57] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "A Head-Mounted Microphone Array for Binaural Rendering," in *International Conference on Immersive and 3D Audio*, Bologna, Italy, 2021, pp. 1–7.

[58] L. Müller, "Perceptual Differences Caused by Altering the Elevation of Early Room Reflections," Master Thesis, Chalmers University of Technology, 2021.

[59] J. Ahrens, *Analytic Methods of Sound Field Synthesis*, 1. Auflage, S. Möller, A. Küpper, and A. Raake, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, p. 307, ISBN: 978-3-642-25742-1. DOI: 10.1007/978-3-642-25743-8.

[60] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "The Far-Field Equatorial Array for Binaural Rendering," in *International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada: IEEE, 2021, pp. 421–425. DOI: 10.1109/ICASSP39728.2021.9414368.

[61] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer," *Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. 805–815, 2021. DOI: 10.1121/10.0005754.

[62] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "Spherical Harmonic Decomposition of a Sound Field Based on Microphones Around the Circumference of a Human Head," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA: IEEE, 2021, pp. 1–5.